Arch Virol (1993) 130: 279–288

Archives of Virology
© Springer-Verlag 1993
Printed in Austria

The gene encoding the nucleocapsid protein: sequence analysis in murine hepatitis virus type 3 and evolution in *Coronaviridae*

D. Décimo[1, 2], **H. Philippe**[3], **Michelle Hadchouel**[1], **M. Tardieu**[1, 2],
and **M. Meunier-Rotival**[1]

[1] INSERM Unité 347 affiliée au CNRS and [2] Laboratoire de Neurovirologie
et Neuroimmunologie, Université Paris XI, Hôpital de Kremlin-Bicêtre
[3] Laboratoire de Biologie Cellulaire 4 (URA CNRS 1134), Université Paris XI, Orsay,
France

Accepted December 15, 1992

Summary. The nucleoprotein-encoding gene (N) of murine hepatitis virus type 3 (MHV 3), from the Mill Hill strain, was cloned and sequenced. It was compared to gene N from other murine coronaviruses and was found to share more similarities with N sequences from MHV 1 and MHV JHM strains than with the published MHV 3 N sequence which is almost identical to MHV A59. We suggest that the evolution of some MHV N sequences resulted from a double recombination phenomenon between two ancestors. Furthermore, comparison of protein N from avian and mammalian coronaviruses leads to the hypothesis that horizontal transfer events of the virus from one host species to another have occurred.

Introduction

Murine hepatitis viruses (MHV) are enveloped RNA viruses belonging to the family *Coronaviridae*. Their genome is a polyadenylated, non-segmented, single stranded RNA [20] coding for three to four major structural proteins according to the antigenic group: the spike glycoprotein (S), the HE antigen, the membrane associated matrix glycoprotein (M), and the internal nucleocapsid protein (N), in addition to non-structural proteins such as an RNA-dependant RNA polymerase [18]. The order of the genes from 5′ to 3′ is: RNA polymerase, HE, S, M, N [9]. During replication, a full-length negative-stranded RNA and six polyadenylated messenger RNAs are synthesized in addition to the full length positive-sense RNA. The subgenomic RNAs form a nested set, all overlapping with the 3′ end of the genomic RNA [10].

Among the MHV group, different strains induce various pathologies [27].

Intraperitoneal injection of MHV 3 from the Mill Hill strain (MHV3 MH) in adult mice results in fulminant hepatitis or in chronic infection of brain [23, 26]. In order to better characterize MHV3 MH and assess its genetic origin, we cloned and sequenced the 3′ end of MHV3 MH RNA, i.e., the nucleoprotein-encoding gene. By comparing nucleotide sequences, we suggest that N genes from murine hepatitis viruses evolved by a double recombination phenomenon between two ancestors. We also report the phylogeny of protein N from avian and mammalian *Coronaviridae*.

Materials and methods

Viruses

MHV3 MH belongs to the family *Coronaviridae*, genus *Coronavirus*, murine hepatitis virus type 3 from the Mill Hill strain. The virus used throughout this work was derived from a single isolate of plaque purification [24] and was stored at − 80 °C.

Cloning of cDNA

Viral particles were purified [21] and the RNA extracted [10]. Viral RNA was used for cDNA synthesis and the cDNA fragments were cloned into the Eco RI site of λ gt 11 phage [28]. MHV 3 recombinant clones were detected by immunodetection of the fusion protein using polyclonal antibodies against MHV3 MH which were prepared by immunizing a rabbit with purified virus [21] and were revealed by a rabbit-PAP system (DAKO) with 3-amino-9-ethylcarbazole as substrate.

DNA and RNA sequencing

Restriction fragments of MHV 3 inserts were cloned in the multisite of M 13 mp 18 and 19 replicative forms. DNA sequencing was performed by the dideoxy chain termination method [15]. RNA sequencing was done [4] on purified virion RNA with a synthetic oligonucleotide (GGCTGATTCCTCCTGCCTC) complementary to positions 146 to 128 (Fig. 1) as primer.

Sequence analyses

Computer analyses were performed using the Bisance facilities [3] on the VAX of CITI 2 in Paris and on PC microcomputers using the MUST package [14]. Amino acid sequences were first aligned using the Clustal program [6] and corrected by hand. For phylogenetic analyses, PAUP version 2.4.1. [22] was used which deduces phylogenetic relations on the basis of maximum parsimony. The following N protein sequences from the Swiss-Prot Database Rel. 23 were used: avian infectious bronchitis virus, IBV (M 21515); turkey enteric coronavirus, TCV (P26020); murine hepatitis viruses: MHV 1 (P18446); MHV3 PM (P18447), MHV A59 (P18448), MHV JHM (P03417); porcine respiratory coronavirus: PRC (P24411), porcine transmissible gastroenteritis viruses: TGEV fs (P05991), TGEV purdue (P04134); bovine coronaviruses: BCV f15 (P19902), BCV mebus (P10527); feline infectious peritonitis virus, FIP (P25909); human coronaviruses: HCV 229E (P15130), HCV OC43 [7].

Results

Sequence of MHV3 MH gene N

The cDNA library constructed from murine hepatitis virus RNA of the Mill Hill strain was screened by immunodetection (data not shown). A positive cDNA was subcloned in M 13 vectors and sequenced. The sequence corresponded to the nucleoprotein-encoding gene but lacked the 5′ end. Viral RNA was then sequenced directly using a synthetic oligonucleotide designed from the cDNA sequence just determined. The resulting composite sequence is shown in Fig. 1 beginning at the initiation codon (accession number X63538 in the EMBL Database).

The sequence of Fig. 1 was translated in the 3 possible reading frames and 2 long open reading frames (ORF) were found. The smaller one (positions 26 to 685) encoded a hypothetical protein of 220 amino acids, 24,009 Da in molecular mass, rather basic (11.4 and 8.2% of basic and acid amino acids, respectively) and serine-rich and leucine-rich (8.6 and 17% of total amino acids, respectively).

The larger ORF (Fig. 1) was 1371 bases long, the 3′ untranslated region contained 295 bases including the first A of the polyA tail, with no canonical polyadenylation signal but containing a 10 base motif (position 1593 to 1602) which is relatively conserved among coronaviruses [16]. The predicted 457 amino acid protein of 50,065 Da is similar to other murine hepatitis nucleocapsid proteins whereas it was 55 kDa in molecular mass when estimated by "Western" blotting (data not shown). The nucleocapsid protein is composed by 14.7% basic amino acids (33 lysine, 30 arginine, and 4 histidine) and 10.3% acidic residues (25 aspartic acid, 22 glutamic acid), giving a basic protein, a property expected for a nucleic acid binding protein. Basic amino acid residues are clustered in the central part of the protein and acidic ones in the carboxy terminus.

Evolution of the nucleoprotein-encoding gene in murine hepatitis viruses

In order to analyze the variability of the nucleoprotein-encoding gene of MHV, all available sequences were aligned (Fig. 1) and compared pairwise (Table 1). The following sequences were used: MHV 1 [13], MHV JHM [17], MHV A59 [1, 13], MHV3 PM [13], and MHV3 MH (this work). The sequence of gene N from MHV S, which was shown to be a recombinant one [13], was not considered. The N sequences from the two MHV 3 strains appeared highly divergent: the gene from MHV3 MH (position 1 to 1371) was longer by 9 bases than the MHV3 PM sequence and was different at 90 positions whereas MHV3 PM and MHV A59 sequences were almost identical (2 bases were different in the coding sequence leading to one different amino acid) as already stated [13]. Such a low level of sequence variation between 2 different strains

```
     M  S  F  V  P  G  Q  E  N  A  G  S  R  S  S  F  G  N  R  A  G  N  G  I  L  K  K  T  T  W  A  D  Q  T  E  R  G  P  N  N    40
3MH ATGTCTTTTGTTCCTGGGCAAGAAAATGCCGGTAGCAGAAGCTCTTTTGGAAACCGCGCTGGTAATGGAATCCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCGCGGACCAAATAAT  120
1                                               C C  T                                                         T
JHM                                             C C                                                               GTT
A59                                  G          C C  T                                                         T
3PM                                  G          C C                                                            T

     Q  N  R  G  R  R  N  Q  P  K  Q  T  A  T  T  Q  P  N  S  G  S  V  V  P  H  Y  S  W  F  S  G  I  T  Q  F  Q  K  G  K  E    80
3MH CAAAATAGAGGCAGGAGGAATCAGCCAAAGCAGACTGCAACTACTCAACCCAATTCCGGGAGTGTGGTTCCCCATTACTCCTGGTTTTCTGGCATTACCCAATTCCAGAAGGGAAAGGAG  240
1                 A                                                                 T        G              T           A
JHM               A A         C                                                     T        G                          A
A59               A                                      C                                               G     A
3PM               A                                      C                                               G     A

     F  K  F  A  D  G  Q  G  V  P  I  A  N  G  I  P  A  S  E  Q  K  G  Y  W  Y  R  H  N  R  R  S  F  K  T  P  D  G  Q  Q  K   120
3MH TTTAAGTTTGCAGATGGACAGGGAGTGCCTATTGCCAATGGAATCCCAGCTTCAGAGCAAAAGGGATATTGGTATAGACACAACCGACGGTCTTTTAAAACACCTGATGGCCAGCAGAAG  360
1      C        C A                       C
JHM    C        C A     A                                 C                   C              T  C
A59    C          A     A                           C                                      C  T                    G
3PM    C          A     A                           C                                      C  T                    G

     Q  L  L  P  R  W  Y  F  Y  Y  L  G  T  G  P  H  A  G  A  E  Y  G  D  D  I  D  G  V  V  W  V  A  S  Q  Q  A  D  T  K  T   160
3MH CAGCTACTGCCCAGATGGTATTTTTACTATCTTGGAACAGGGCCCCATGCTGGCGCAGAGTATGGCGACGATATCGACGGAGTTGTCTGGGTCGCAAGCCAACAGGCCGACACTAAGACC  480
1
JHM                                              T                               A                                G    G
A59   AT                               C                 A  CAGT     A   AGC  T  A  T  CT       T    A AGC  A  G     C  T
3PM   AT                               C                 A  CAGT     A   AGC  T  A     CT       T    A AGC  A  G     C  T

     T  A  D  I  V  E  R  D  P  S  S  H  E  A  I  P  T  R  F  A  P  G  T  V  L  P  Q  G  F  Y  V  E  G  S  G  R  S  A  P  A   200
3MH ACTGCCGATATTGTTGAAAGGGACCCAAGTAGCCATGAGGCTATTCCTACTAGGTTTGCGCCCGGTACGGTATTGCCTCAAGGTTTTTATGTTGAAGGCTCAGGAAGGTCTGCACCTGCT  600
1
JHM T
A59 CGCT T        C              C  T                                C              G  C                 T
3PM CGCT T        C              C  T                                C              G  C                 T

     S  R  S  G  S  R  S  Q  S  R  G  P  N  N  R  S  R  S  S  S  N  Q  R  Q  P  A  S  T  V  K  P  D  M  A  E  E  I  A  A  L   240
3MH AGTCGATCTGGTTCGCGGTCACAATCCCGTGGGCCAAATAATCGCTCTAGAAGCAGCTCCAACCAGCGCCAGCCTGCCTCTACTGTAAAACCTGATATGGCCGAAGAAATTGCTGCTCTT  720
1                                                G
JHM                   C                          G          T
A59   C                                          G          T
3PM   C                                          G          T

     V  L  A  K  L  G  K  D  A  G  Q  P  K  Q  V  T  K  Q  S  A  K  E  V  R  Q  K  I  L  N  K  P  R  Q  K  R  T  P  N  K  Q   280
3MH GTTTTGGCTAAGCTCGGTAAAGATGCCGGCCAGCCCAAGCAAGTAACAAAGCAAAGCGCCAAAGAAGTCAGGCAGAAAATTTTAAACAAGCCTCGTCAAAAGAGGACTCCAAACAAGCAG  840
1
JHM                                    T                    T
A59                                                G        T                                      C
3PM                                                G        T                                      C

     C  P  V  Q  Q  C  F  G  K  R  G  P  N  Q  N  F  G  G  P  E  M  L  K  L  G  T  S  D  P  Q  F  P  I  L  A  E  L  A  P  T   320
3MH TGCCCTGTGCAGCAGTGTTTTGGAAAGAGAGGCCCCAATCAAAATTTTGGAGGCCCTGAAATGTTAAAACTTGGAACTAGTGATCCGCAGTTCCCCATTCTTGCAGAGTTGGCCCCAACC  960
1                                            G            T                               A                                A
JHM      A                                   G                                            A                                A
A59      A                                   G            T                               A                          T     A
3PM      A                                   G            T                               A                          T     A

     P  S  A  F  F  F  G  S  K  L  E  L  V  K  K  N  S  G  G  A  D  E  P  T  K  D  V  Y  E  L  Q  Y  S  G  A  V  R  F  D  S   360
3MH CCTAGTGCCTTCTTCTTTGGATCTAAATTAGAATTGGTCAAAAAGAACTCTGGTGGTGCTGATGAACCCACCAAAGATGTGTATGAGCTGCAATATTCAGGTGCAGTTAGATTTGATAGT 1080
1                                                                    C                    AT    G            A
JHM G  G                                                            G
A59 GT G                                           T
3PM GT G                                           T

     T  L  P  G  F  E  T  I  M  K  V  L  N  E  N  L  N  A  Y  Q  D  Q  A  G  G  A  D  V  V  S  P  K  P  Q  R  K  R  G  Q  R   400
3MH ACTCTACCTGGTTTTGAGACTATCATGAAAGTGTTGAATGAGAATTTGAACGCCTACCAGGATCAAGCTGGTGGTGCAGATGTAGTGAGCCCCAAGCCCCAAAGAAAGAGAGGGCAAAGA 1200
1        C  A  A                                    G T                                         A                     AC  A
JHM                                                    T         A      A                        T     T  G            AC  AG
A59                                                    T         A ---G A               G        A              ---     GT
3PM                                                    T         A ---G A               G        A              ---     GT

     Q  V  A  Q  K  K  N  D  E  V  D  N  V  S  V  A  K  P  K  S  S  V  Q  R  N  V  S  R  E  L  T  P  E  D  R  S  L  L  A  Q   440
3MH CAGGTGGCTCAAAAGAAGAATGATGAAGTAGATAATGTAAGCGTTGCAAAGCCCAAAAGCTCTGTGCAGCGAAATGTAAGTAGAGAATTAACCCCAGAGGATAGAAGTCTGTTGGCTCAG 1320
1     AAA     ------T   A G                                                                         T      C T
JHM   AAA     ------    A                                                                  G        T      C C  C  TC
A59   ---      GG A     A
3PM   ---      GG A     A

     I  L  D  D  G  V  V  P  D  G  L  E  D  D  S  N  V                                                                      457
3MH ATCCTTGATGATGGCGTAGTGCCAGATGGGTTAGAAGATGACTCTAATGTGTAAAGAGAATGAATCCTATGTCGGCACTCGGTGGTAACCCCTCGCGAGAAAGTCGGGATAGGACACTCT 1440
1                    T     T
JHM      A
A59                                                                             G
3PM                                                                             G
3MH CTATCAGAATGGATGTCTTGCTGTCATAACAGATAGAGAAGGTTGTGGCAGACCCTGTATCAATTAGTTGAAAGAGATTGCAAAATAGAGAATGTGTGAGAGAAGTT-GCAAGGTCCTAC 1559
1                                                                                                              A
JHM                                                                                                            A
A59                                                                                                            A
3PM                                                                                                            A
3MH GTCTAACCATAAGAACGGCGATAGGCGCCCCC-TGGGAAGAGCTCACATCAGGGTACTATTCCTGCAATGCCCTAGTAAATGAATGAAGTTGATCATGGCCAATTGGAAGAATCACA    1675
1                                   C
JHM                                 -
A59                                 -                             T
3PM                                 -                                                                                 G
```

Table 1. Nucleoprotein-encoding gene sequence similarity among the different MHV strains (%)

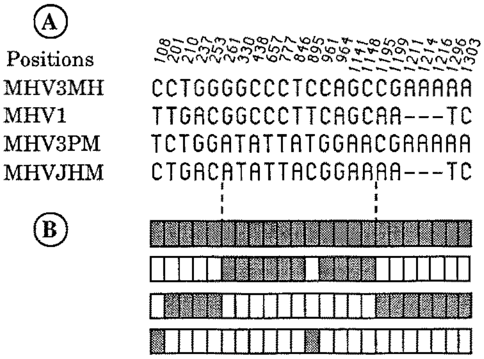
	3 MH	1	JHM	A 59	3 PM
3 MH		96.9 (96.4)	96.1 (95.3)	94.4 (93.4)	94.5 (93.5)
1	96.3		96.8 (96.2)	93.7 (92.5)	93.7 (92.5)
JHM	94.5	95.6		93.7 (92.5)	93.7 (92.6)
A 59	94.3	93.8	93.4		99.8 (99.8)
3 PM	94.1	93.5	93.4	99.8	

Numbers above the diagonal represent the per cent similarities among nucleotides (from the initiation codon to the first A of the polyA tail; in parentheses: coding sequence). Numbers below the diagonal are per cent similarities at the amino acid level

is within experimental error so that MHV3 PM and MHV A59 nucleocapsid sequences were considered to be the same one.

We then turned to an analysis of relationships between strains by identifying informative positions. Informative positions for the parsimony analysis are defined as those at which 2 different nucleotides are present at least twice each. With 4 sequences (without MHV A59), there are 23 informative positions, spread over the whole sequence (Fig. 2 A). A graphical representation of the informative positions and of their character state is provided in Fig. 2 B. This juxtaposition of the nucleotides in the informative positions can be interpreted in terms of associations. The informative positions at the two ends of the sequence group MHV 1 with MHV JHM while those in the central region group MHV 1 with MHV3 MH.

Fig. 1. Nucleotide sequence of the nucleocapsid protein-encoding gene from MHV3 MH and its deduced amino acid sequence. Alignment of nucleotide sequences from the MHV nucleoprotein-encoding gene. Numbering is relative to the A residue of the start codon of MHV3 MH and amino acid translation shown above, using the single letter code, refers to MHV3 MH. The nucleotide sequence was determined from a cDNA clone, from positions 55 to 1675 and on viral RNA, from 1 to 100, using a synthetic oligonucleotide (complementary to positions 146 to 128) as primer. MHV sequences are compared to the longest one: MHV3 MH. Gaps are indicated by a dash and blanks indicate identical nucleotides



```
Positions  108 201 210 237 253 261 330 438 657 777 846 895 961 964 1141 1148 1195 1199 1211 1214 1216 1296 1303
MHV3MH     CCTGGGGCCCTCCAGCCGAAAAA
MHV1       TTGACGGCCCTTCAGCAA---TC
MHV3PM     TCTGGATATTATGGAACGAAAAA
MHVJHM     CTGACATATTACGGAAAA---TC
```

Fig. 2. Informative positions of four MHV N sequences. **A** Informative positions (two different bases at a given position, present at least twice each) are listed in the order (upper line) by which they appear in the sequences of Fig. 1. Gaps are indicated by a dash. Vertical lines indicate the limits of the 3 domains. **B** Schematic representation. Blank and shaded boxes indicate identity of nucleotides, respectively, when read vertically. By convention, the nucleotides of the first sequence were all shaded without any implication as to whether the nucleotide association is parental within this sequence

Evolution of coronavirus protein N in birds and mammals

We compared all N protein sequences of avian and mammalian coronaviruses available in the Swiss-Prot Database Rel. 23 plus HCV OC43 [7]. Amino acid sequences were first aligned using the Clustal program [6]; 222 positions can be aligned unambiguously. In a first step, we verified by PAUP [22] that the N sequences of viruses infecting the same species (mouse, pig, and cattle) were clustered together on the tree. We chose one sequence for pig and cattle viruses and two for MHV (MHV 1 and 3), the most distant ones. For avian and turkey viruses and among human viruses which are not closely related on the tree, individual sequences were used. The phylogenetic tree of the various N proteins constructed by the parsimony method is shown in Fig. 3 A, while the cladogram of birds and mammals based on morphological characters [12] is provided in Fig. 3 B. First, we found that human OC43, bovine, and turkey enteric coronaviruses give a monophyletic group. The distances between these 3 viruses were smaller than between MHVs. Second, porcine and feline viruses were more closely related than their respective hosts.

The same analysis was performed on protein M sequences of the same coronaviruses (when available) and an identical most parsimonious tree was obtained (data not shown).

Discussion

We sequenced the 3′ part of the MHV3 MH genome and showed that it contained 2 open reading frames. The smallest one was also detected in other MHV N sequences [13], although these ORF are shorter (207 aa). The only exception is for the MHV JHM sequence in which a premature stop codon is found. This

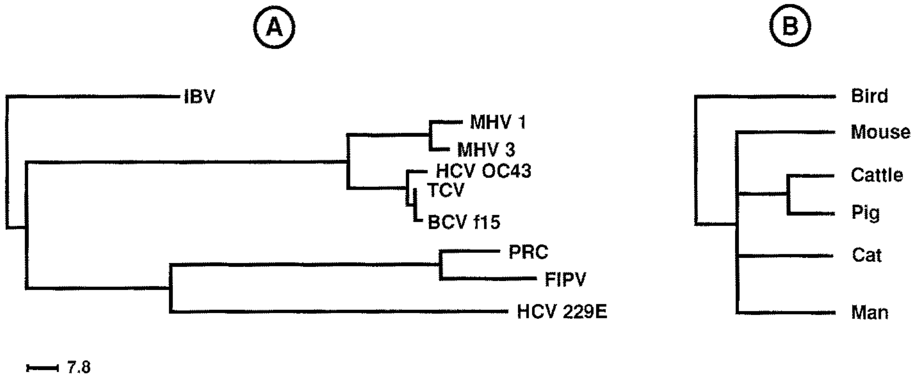


Fig. 3. Phylogenetic tree of conserved regions of nucleoprotein N among *Coronaviridae* (**A**) and cladogram of the hosts (**B**). **A** The N proteins used to build the tree are listed in Materials and methods. 222 positions (67–217 and 270–328 on MHV 1) could be aligned unambiguously, among which 134 were informative for the parsimony. The scale bar represents 7.8 units defined as the number of substitutions estimated by the PAUP program [22]. All positions are equally weighted, unordered, multistate characters. Gaps are used as a character state in the analysis. The number of steps is equal to 421 and the consistency index to 0.96. There is an equiparsimonious tree which places HCV OC43 as sister group of the turkey coronavirus. **B** The recently published cladogram of mammals [12] displays a multifurcation of rodents, carnivores, artiodactyls, and primates

hypothetical protein has not yet been detected. The longer ORF which codes for the nucleocapsid protein presents a discrepancy between the calculated and estimated molecular mass which could be explained by post-translational modifications such as phosphorylation of serine residues, which account for 8.5% of total amino acids. Phosphorylation of MHV nucleoproteins has been described [19].

We showed that the nucleotide sequences of gene N of 2 viruses which were classified under the name MHV 3 are different. The N gene of the virus used in our laboratory (Mill Hill strain) seemed original by comparison with the MHV3 PM [13] which is almost identical to MHV A59. The identity of these 2 sequences could be explained by genetic recombination between the 3′ termini of MHV 3 and A 59 genomes leading to a common N gene. Such recombination processes have been observed among laboratory strains of murine coronaviruses, under special laboratory conditions [5, 8, 11]. Alternative hypotheses can also be considered such as a mixture of viruses in the infected animal from which MHV 3 was first isolated or a mixture of the two strains in the cultured cells.

A double recombination event between two ancestors is suggested by the distribution of the informative positions in three regions, two of which are of the same kind (MHV 1 is the same as MHV JHM at the ends and MHV 1 is the same as MHV3 MH in the central part). Only 2 positions (108 and 895) support grouping MHV 1 with MHV3 PM in the left and central portion of the sequence, respectively, but this can be interpreted as convergence in 2 out of 23 positions which does not blur the phylogenetic information. It is not possible to determine the polarity of recombination or the nature of the ancestors.

The junctions of the putative recombination events would be between positions 254 and 260 and between positions 1149–1194, as indicated by the dashed lines in Fig. 2. It is interesting to note that a three domain structure for the nucleoprotein has been suggested, with spacers A (positions 420 to 486) and B (positions 1143 to 1215) defined as hot spots of variability between domains [13]. Spacer B corresponds to the second putative recombination region that we have identified, but spacer A is not located within the first recombination region. If instead of comparing nucleotide sequences, we converted the sequence into amino acids, very little information appears since most nucleotide informative positions were silent at the amino acid level (only 6 informative positions remain at the amino acid level).

As coronaviruses were shown to have a marked host specificity [9, 20], it is generally assumed that the virus and its host evolved in parallel, the position of the former on the phylogenetic tree mimicking the position of the latter. We checked this hypothetical co-evolution between virus and host by constructing the phylogenetic tree of M and N protein sequences. The M and N genes are located next to each other and it would have been better to use 2 distant genes, one at the beginning of the genomic RNA, the N gene being located at the 3′ end, but the sequences are not all available. Nevertheless, the fact that we obtain the same tree with the sequences of 2 different viral proteins is a strong argument in favor of the generality of the tree.

The N sequences of turkey, bovine, and human OC43 coronaviruses are so closely related [25] that they are less polymorphic than the MHV N sequences, leading to the conclusion that the interspecific diversity is smaller than the intraspecific polymorphism. The HE proteins of BCV and HCV OC43 were also shown to be very similar [29]. It is possible that very recent host changes have occurred as a result either of domestication or of some laboratory forced infections.

The phylogenetic tree cannot be explained without one horizontal transfer of viruses between feline and porcine hosts on a large evolutionary scale (over million years) because it is well known that Artiodactyla (pig and cattle) are not closely related to rodents or carnivores [2, 12] (see Fig. 3 B). Thus, we could infer from the phylogenetic tree of N proteins that very recent transfer phenomena have occurred in parallel with horizontal transfer and co-evolution.

In conclusion, our analyses suggest that co-evolution between the virus and its host and horizontal transfer both participate in the evolution of *Coronaviridae*. The relative importance of the 2 phenomena could be studied if more "wild" viral strains were available from "wild" hosts.

Acknowledgements

We thank Dr. H. Laude for helpful advices, Dr. A. Adoutte for helpful discussions, and Dr. L. Sperling for critically reading the manuscript. D. D. was supported by SESEP.

References

1. Armstrong J, Smeekens S, Rottier P (1983) Sequence of the nucleocapsid gene from murine coronavirus MHV A59. Nucleic Acids Res 11: 883–891
2. Carroll RL (1988) Vertebrate palaeontology and evolution. WH Freeman, New York
3. Dessen P, Fondrat C, Valencien C, Mugnier C (1990) Bisance: a French service for access to biomolecular sequence databases. Comput Appl Biosci 6: 355–356
4. Fichot O, Girard M (1990) An improved method for sequencing of RNA templates. Nucleic Acids Res 18: 162
5. Fu K, Baric RS (1992) Evidence for variable rates of recombination in the MHV genome. Virology 189: 88–102
6. Higgins DG, Sharp PM (1988) Clustal: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244
7. Kamahora T, Soe LH, Lai MMC (1989) Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. Virus Res 12: 1–9
8. Keck JG, Soe LH, Makino S, Stohlman SA, Lai MMC (1988) RNA recombination of murine coronaviruses: recombination between fusion-positive hepatitis virus A 59 and fusion-negative mouse hepatitis virus 2. J Virol 62: 1989–1998
9. Lai MMC (1990) Coronaviruses: organization, replication, and expression of genome. Annu Rev Microbiol 44: 303–333
10. Lai MMC, Brayton PR, Armen RC, Patton CD, Pugh C, Stohlman SA (1981) Mouse hepatitis virus A 59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J Virol 39: 823–834
11. Makino S, Keck JG, Stohlman SA, Lai MMC (1986) High-frequency RNA recombination of murine coronaviruses. J Virol 57: 729–737
12. Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356: 121–125
13. Parker MM, Masters PS (1990) Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. Virology 179: 463–468
14. Philippe H (1992) Management utilitarians for sequences and trees. University of Paris-Sud, Orsay
15. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with high chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
16. Schreiber SS, Kamahora T, Lai MMC (1989) Sequence analysis of the nucleocapsid protein gene of human coronavirus 229 E. Virology 169: 142–151
17. Skinner MA, Siddell SG (1983) Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. Nucleic Acids Res 11: 5045–5054
18. Spaan W, Cavanagh D, Horzinek MC (1988) Coronaviruses: structure and genome expression. J Gen Virol 69: 2939–2952
19. Stohlman SA, Lai MMC (1979) Phosphoproteins of murine hepatitis viruses. J Virol 32: 672–675
20. Sturman LS, Holmes KV (1983) The molecular biology of coronaviruses. Adv Virus Res 28: 35–112
21. Sturman LS, Holmes KV, Behnke J (1980) Isolation of coronavirus envelope glycoproteins and interaction with viral nucleocapsid. J Virol 33: 449–462
22. Swofford DL (1985) PAUP: phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign
23. Tardieu M, Boespflug O, Barbé T (1986) Selective tropism of a neurotropic coronavirus for ependymal cells, neurons, and meningeal cells. J Virol 60: 574–582
24. Tardieu M, Goffinet A, Harmant-van Rijckevorsel G, Lyon G (1982) Ependymitis, leukoencephalitis, hydrocephalus, and thrombotic vasculitis following chronic infection by mouse hepatitis virus 3 (MHV 3). Acta Neuropathol 58: 168–176

25. Verbeek A, Tijssen P (1991) Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus. J Gen Virol 72: 1659–1666
26. Virelizier JL, Virelizier AM, Allison AC (1975) Neuropathological effects of persistent infection of mice by mouse hepatitis virus. Infect Immunol 12: 1127–1140
27. Wege H, Siddell S, Ter Meulen V (1982) The biology and pathogenesis of coronaviruses. Curr Top Microbiol Immunol 99: 165–200
28. Young RA, Davies RW (1983) Efficient isolation of genes by using antibody probes. Proc Natl Acad Sci USA 80: 1194–1198
29. Zhang X, Kousoulas KG, Storz J (1992) The haemagglutinin/esterase gene of human coronavirus strain OC 43: phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. Virology 186: 318–323

Authors' address: M. Meunier-Rotival, INSERM U 347, 80 rue du Général Leclerc, F-94276 Le Kremlin-Bicêtre Cedex, France.

Received July 21, 1992